frontiers | Frontiers in Virtual Reality

TYPE Original Research
PUBLISHED 23 June 2025
DOI 10.3389/frvir.2025.1574965

OPEN ACCESS

EDITED BY
Andrea Sanna,
Polytechnic University of Turin, Italy

REVIEWED BY
Etienne Peillard,
IMT Atlantique Bretagne-Pays de la Loire, France
Unais Sait,
Free University of Bozen-Bolzano, Italy

*CORRESPONDENCE
Catarina G. Fidalgo,
cfidalgo@andrew.cmu.edu

RECEIVED 11 February 2025
ACCEPTED 12 May 2025
PUBLISHED 23 June 2025

CITATION
Fidalgo CG, Yan Y, Sousa M, Jorge J and Lindlbauer D (2025) Exploring AR hand augmentations as error feedback mechanisms for enhancing gesture-based tutorials. *Front. Virtual Real.* 6:1574965. doi: 10.3389/frvir.2025.1574965

Exploring AR hand augmentations as error feedback mechanisms for enhancing gesture-based tutorials

Catarina G. Fidalgo[1,2,3]*, Yukang Yan[1], Mauricio Sousa[4], Joaquim Jorge[2,3] and David Lindlbauer[1]

[1]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, United States, [2]Computer Science and Engineering Department, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal, [3]Graphics and Interaction Group, INESC-ID, Lisboa, Portugal, [4]Department of Computer Science, University of Toronto, Toronto, ON, United States

Self-guided tutorials from videos help users learn new skills and complete tasks with varying complexity, from repairing a gadget to learning how to play an instrument. However, users may struggle to interpret 3D movements and gestures from 2D representations due to different viewpoints, occlusions, and depth perception. Augmented Reality (AR) can alleviate this challenge by enabling users to view complex instructions in their 3D space. However, most approaches only provide feedback if a live expert is present and do not consider self-guided tutorials. Our work explores virtual hand augmentations as automatic feedback mechanisms to enhance self-guided, gesture-based AR tutorials. We evaluated different error feedback designs and hand placement strategies on speed, accuracy and preference in a user study with 18 participants. Specifically, we investigate two visual feedback styles — *color feedback*, which changes the color of the hands' joints to signal pose correctness, and *shape feedback*, which exaggerates fingers length to guide correction — as well as two placement strategies: *superimposed*, where the feedback hand overlaps the user's own, and *adjacent*, where it appears beside the user's hand. Results show significantly faster replication time when users are provided with color or baseline no explicit feedback, when compared to shape manipulation feedback. Furthermore, despite users' preferences for adjacent placement for the feedback representation, superimposed placement significantly reduces replication time. We found no effects on accuracy for short-time recall, suggesting that while these factors may influence task efficiency, they may not strongly affect overall task proficiency.

KEYWORDS

tutorials, training, augmented reality, hand gestures, error feedback, virtual hand augmentations

1 Introduction

Online video tutorials are ubiquitous resources for people to learn new skills and tackle tasks of varying complexity (de Koning et al., 2018; Mayer et al., 2020), from crocheting to learning to play instruments. Many tutorials are filmed from the point of view of the person demonstrating the task (Li et al., 2023). This perspective enables viewers to see specific hand poses (e.g., sign language), hand movements (e.g., chopping an onion), or how to

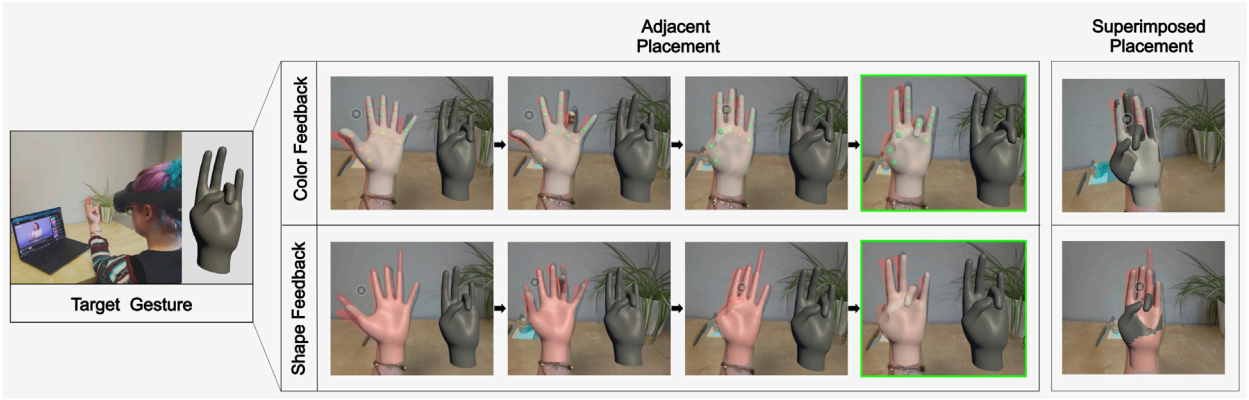

FIGURE 1
Augmented reality view of a user replicating a hand gesture from a Youtube video. On the top, the user's virtual hand shows Color Feedback: joint spheres are shaded along a red-to-green gradient depending on how closely each joint matches the target. Red indicates large error, green indicates near-perfect alignment. On the bottom, the hand shows Shape Feedback: finger segments are exaggerated in length based on error magnitude, forming visual "error bars." Both feedback types are shown in adjacent (beside user's hand) and superimposed (overlaid on user's hand) placements.

manipulate an object (e.g., device assembly). However, when watching such tutorials on the 2D displays of desktop computers or smartphones, it is difficult for users to interpret and mimic the hand movements, gestures, and poses naturally performed in 3D. This mismatch arises from various challenges, for example, estimating depth and motion paths and speeds, dealing with viewpoints that differ from the one in the video, or general occlusion (Mohr et al., 2017).

Augmented Reality (AR) has the potential to alleviate this challenge by enabling users to view complex instructions in a 3D space. Showing users a 3D representation of the actions they have to perform facilitates their understanding of depth and spatial relationships (Krolovitsch and Nilsson, 2009), and provides a more immersive experience (Brunet and Andújar, 2015). However, providing only a 3D representation (e.g., showing a hand pose to learn sign language) can be insufficient for users to accurately mimic an action or learn a task. Without real-time feedback (Herbert et al., 2018), users may not realize whether they acted accurately, which impairs their performance.

To address this challenge, we explore how to best provide real-time feedback for AR tutorials that involve complex hand poses and gestures. Real-time feedback aims to increase users' hand pose accuracy (e.g., reducing joint offsets), memorization, and learning. Prior work provided users with live feedback that guided them through tasks by displaying an additional pair of hands in AR (Amores et al., 2015; Sodhi et al., 2013; Tecchia et al., 2012). Those approaches, however, rely on live feedback from a remote helper, limiting the availability of this type of feedback. Our work explores ways to guide users without requiring additional (expert) users to deliver the guidance. We provide users who aim to mimic and recall complex hand movements with different types of real-time feedback on their accuracy. We hypothesize that visually displaying errors directly on the source (i.e., the hand) will guide the user's attention to the task (Ozcelik et al., 2010), reduce diverted attention between instruction and task, and improve retention (Jamet et al., 2008).

We test different parameters for AR hand augmentations for error feedback by identifying key modification dimensions based on prior work. Specifically, we investigate two *feedback styles* designed to convey performance errors, described in Section 3: (1) *Color feedback*, where joint spheres change color to indicate correctness or incorrectness of finger poses, and (2) *Shape feedback*, a novel condition where the shape of the fingers is manipulated to exaggerate the correct pose and guide adjustment (See Figures 1, 3). Additionally, we explore the optimal *placement* of these augmented hands to guide users, exploring both direct and indirect methods — either *superimposed*, directly over the user's own hands, or *adjacent*, displayed beside the user's hands.

We evaluated the effectiveness and usability of different feedback styles and placements for hand augmentations in a user study with 18 participants. Participants were asked to perform hand gestures from Portuguese Sign Language and Mudras, hand gestures used in a traditional Indian dance. Participants were asked to replicate the gestures as quickly as possible during one task, and as accurately as possible during another task. Results show significantly faster replication time when users are provided with color feedback or no explicit feedback compared to shape manipulation feedback, as well as faster times for superimposed placement. Participants did, however, prefer adjacent placement of the augmentations. In summary, we contribute:

- A set of novel hand augmentations for self-guided AR tutorials, exploring two feedback styles (color and shape-based) and two placement strategies (superimposed and adjacent) for conveying performance errors;
- Insights from a user study ($N = 18$) assessing the impact of different hand augmentation dimensions in gesture replication speed, accuracy and user preferences.

2 Related work

AR has been shown to have a positive impact for instructional guidance (Fidalgo et al., 2023) and immersive learning

environments in terms of task performance, reducing workload (Tang et al., 2003), helping with structural perception (Gupta et al., 2012), and preventing the systematic mislearning of content (Büttner et al., 2020), providing similar learning outcomes compared to video-based tutorials but with higher usability and satisfaction (Morillo et al., 2020). In the following, we discuss related work that aims to address the challenges of creating effective AR tutorials.

2.1 Designing AR tutorials

Previous work has investigated how to effectively create AR content and develop the appropriate tools for AR applications. Huang et al. (2021) introduced an adaptive task tutoring system for machine operations that enables experts to record machine task tutorials via embodied demonstration. Their system adapted the displayed AR content, adjusting the number of interface elements based on user behavior. Liu Z. et al. (2023) proposed InstruMentAR, a system that automates AR tutorial generation by automatically recording user demonstrations and generating AR visualizations accordingly. Their multi-modal approach provides haptic feedback when a user is performing or is about to perform a mistake.

Such systems rely on the creation of new content specific to AR. Others have explored leveraging existing 2D video demonstrations for synthesizing AR tutorials. Yamaguchi et al. (2020) created step-by-step animations from video-based assembly instructions, enabling users to see the extracted instructions overlaid onto the current workpiece using an AR "magic mirror" setup. Similarly, Stanescu et al. (2023) captured information about part geometry, assembly sequences, and action videos from RGB-D cameras, recorded during user demonstrations. They provide real-time spatially-registered AR hints for each object part.

Besides assembly tasks, others have explored tutorials for tasks such as painting, soldering, makeup, and decorating. Mohr et al. (2017) extracted motion information from videos and registered these in the user's real-world object. Goto et al. (2010) and Langlotz et al. (2012) leverage existing resources by rendering instructional videos in AR but adapting the display position to the user's viewpoint. Jo et al. (2023) also explored different layouts to display instructional videos. They found that dynamic layouts generally led to fewer timing and posture errors, and that head movement during screen-based monitoring decreases performance. Dürr et al. (2020) suggest that visualizations using continuously moving guidance techniques achieve higher movement accuracy with realistic shapes. Zagermann et al. (2017) also showed that the impact of input modality and display size on spatial memory is not straightforward, but characterized by trade-offs between spatial memory, efficiency, and user satisfaction.

Besides video instructions, Rajaram and Nebeling (2022) showed that enhancing paper-based AR interactions can benefit learning and support students' diverse learning styles. Nonetheless, online content, including text descriptions and video tutorials, usually requires existing knowledge to be understood (Skreinig et al., 2022). Most current AR-based tutorial systems do not provide real-time error feedback to users. This makes judging whether certain tasks are performed correctly or where errors occur is challenging. Our work aims to provide guidelines on delivering such feedback to users for AR tutorials.

2.2 Gesture-based AR tutorials

Previous research explored communicating body movement over distance for collaboration, assistance, training and guidance. We focus on hand gestures given their practical and cultural importance (Flanagan and Johansson, 2002).

Tecchia et al. (2012) used depth sensors to present local users with gestural instructions from a remote expert in VR. Sodhi et al. (2012) combined a low-cost depth camera and a projector to display visual cues directly onto a user's body. Their approach enhanced user's understanding and execution of movements, when compared to animation videos on a computer screen. BeThere (Sodhi et al., 2013) used mobile AR to render the remote participant's hand in the local person's environment. These works rely on having a real-time expert demonstrating the task and providing feedback, whereas we aim for autonomous task guidance.

In the context of instrument learning, Torres and Figueroa (2018) used 2D markers to render spatially annotated 2D cues on a guitar, while Skreinig et al. (2022) generated interactive AR guitar tutorials from tablatures. They captured user input by comparing the emmited *versus* the expected sound while playing a chord, visually highlighting the error region when a mistake occurred. They found that highlighting the regions of importance on the fretboard helped users understand finger placement better. Liu R. et al. (2023) explored how to optimize body posture in piano learning by superimposing hand postures of a pre-recorded teacher over the learner's hands. They evaluated the differences between the recorded (student) and tutorial (teacher) movements through discrepancy metrics. A pilot study suggested that discrepancy displays result in more correct practicing of finger-refined movements, with users preferring having motion overlay on a single keyboard rather than separate keyboards. Zhou et al. (2022) compared visual guidance from an MR mirror and a humanoid virtual instructor with traditional screen-based movement guidance. They found that seeing an overlaid body offers better acquisition performance and a stronger sense of embodiment for upper-body movement than traditional 2D screen-based guidance. Lilija et al. (2021) embedded guidance directly into the user's avatar to minimize visual distraction, instead of relying on external cues such as arrows. Through two experiments, they demonstrated that this technique improves the short-term retention of target movement. Our work explores how similar parameters (e.g., overlay vs. external cues, feedback type) affect hand gesture-based tutorials.

Other work investigated how human dexterity is affected by different hand visualization methods in VR. Voisard et al. (2023) showed that hand visualizations with varying opacity influence the motor dexterity of participants when they perform a task that requires fine hand movements. They demonstrate the potential advantages of less obstructive hand visualizations. Conversely, Ricca et al. (2020), Ricca et al. (2021) showed that, although users prefer to have a visual representation of their hands in VR, they achieved similar and correlated performance without hand visualizations for a tool-based motor task in VR. This showcases the complexity of efficient representations of hands in the context of guidance, depending on level of user expertise (cf. Knierim et al., 2018), task type and objective.

Wang et al. (2024) identified four types of essential information for visual guidance in this context of enhancing precise hand

interaction in VR, including *error* (What's wrong?), *target* (What is correct?), *direction* (What way is it?), and *difference* (How far is it?). Similarly, Yu et al. (2024) devised a design space for corrective feedback focused on *level of indirection* (i.e., disparities between current movement and target), feedback *temporality* and *presentation*, *information level* and *placement*. We build on these works and further investigate some of these dimensions in the context of precise static gestures, specifically how hand feedback influences task performance in AR, i.e., when people simultaneously see their hand and the guidance.

3 Hand augmentations

We explore hand augmentations as a form of feedback mechanism for gesture-based tutorials. All feedback types and placements are depicted in Figure 1.

We aim to enable users to autonomously use the tutorials to build their skills without needing other users or experts. Following Endow and Torres (2021), we believe that tutorials using media such as AR facilitate users' learning process.

We consider the factors *feedback type* (color, shape) and *placement* for the augmentations. We chose these parameters based on previous research on information visualization and AR to convey performance and errors. Note that we do not see those as an exhaustive list of possible parameters and plan to explore a larger design space in the future, such as textual hints and arrows (Oshita et al., 2019), transparency levels (Barioni et al., 2019), rubber band like augmentations (Yu et al., 2020) and trajectories (Clarke et al., 2020), as suggested by Diller et al. (2024) in their survey on visual cue based corrective feedback for motor skill training in MR. Additionally, while we believe that our approach generalizes to dynamic hand gestures, we hope to explore this aspect in the future. We refer to the representation that shows the feedback as *target hands*.

3.1 Feedback style

We explore providing feedback on users' gestures in two different styles, leveraging *Color* or *Shape*.

3.1.1 Color

We visualize the angular error for each joint through color, i.e., the hand's joints change color from green to red depending on the accuracy of the joint position compared to a target gesture. In other words, *Color Feedback* on the joints serves as a heatmap for accuracy. We chose colors for their natural associations and psychological effects. Jacobs and Suess (1975), and Spielberger (1970) showed that higher state-anxiety is more associated with yellow and red tones, compared to blue or green. Additionally, we see the use of reds and greens associated with particular connotations. For example, in many software interfaces and applications, the color green is commonly used to signify success, correctness, or a positive outcome, while red indicates failure, error, or a negative outcome. Similarly, in heatmaps, green is usually associated with lower magnitudes while red is associated with higher magnitudes.

3.1.2 Shape

We provide users with *Shape Feedback* to indicate the accuracy of individual gestures. The user's fingers change in size to reflect error magnitude. Effectively, each finger acts as an error bar, elongating based on the magnitude of the error over that finger's joints. This technique is inspired by the work of Abtahi et al. (2022) on Beyond Real Interactions, as well as distant reaching techniques such as Go-Go (Poupyrev et al., 1996). Furthermore, it is inspired by work of McIntosh et al. (2020), who scaled different parts of the avatar's body, including arms and fingers, to adapt to different tasks. These approaches demonstrate the viability of using non-literal augmentations for guiding motor tasks and user attention in immersive environments. While we acknowledge that shape feedback is a more abstract and novel design, we sought to explore how such expressive techniques might support or hinder performance in gesture learning tasks.

3.2 Placement

Besides the parameters of *in-situ* feedback, we explore where to best place the target hands, i.e., the hands that display the feedback. Previous work (Feuchtner and Müller, 2018; Liu R. et al., 2023; Schjerlund et al., 2021) underscored the value of effective visual placement. We explore two different placement strategies: 1) adjacent and 2) superimposed.

The *adjacent placement*, represents an indirect mapping, where the hands demonstrating a gesture (*target hands*) are shifted to the side of the user's hands, accompanying the rotation and translation of the movement. This configuration is analogous to following a demonstrator nearby and may involve additional spatial interpretation.

In contrast, the *superimposed placement* provides a direct mapping, where the target hands are rendered directly on top of the user's own hands. This approach aims to minimize the need for mental transformation, potentially making interactions more intuitive by visually aligning the intended gesture with the user's own motion path.

Pilot tests helped refine both feedback styles, especially the transformation parameters as described in section 3.3, to ensure that people could visually interpret and decode the exaggerated shape cues or colors during interaction.

3.3 Implementation

Users are presented with virtual hand gesture tutorials in AR through a Meta Quest Pro headset. Our prototype software is created using Unity3D. We first compute the error between users' hand pose and the target. We track the user's hands in real-time using the integrated tracking from the Quest Pro and compare the tracked pose with the target hands displayed in the tutorial. To assess the accuracy of the user's hand movements, we calculate the angular difference between each hand joint and its corresponding position in the target gestures (see Figure 2).

This is calculated as the normalized dot product between the quaternions representing each of the user's hand joints and those representing the target gestures. This feedback is delivered using one of the hand augmentation styles previously discussed.

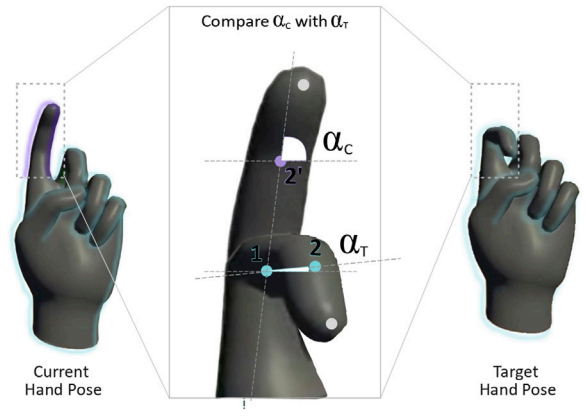

FIGURE 2
Diagram comparing current and target hand poses. On the left, a simplified hand model with an extended index finger represents the current gesture. On the right, a similar hand model shows the index finger arched downward, representing the target pose. Centered between the two hands, a zoom-in detail focuses on the index finger joint, visually superimposing both current and target finger orientations. Angular differences between the two poses are marked with labeled arcs, alfa C for the current joint angle and alfa T for the target, showing the angular difference due to misalignment.

For color feedback, the angular error drives a color interpolation between green and red. We consider the joint to be in a correct position when the difference in angle to the corresponding joint in the target gesture is smaller than 2° and incorrect when it is larger than 60°. We set intermediate thresholds between these extremes (yellow at 8° and orange at 20°) to create a smoother color variation.

For shape feedback, we first compute an average error magnitude over that finger's joints, in degrees, and distribute this over each finger segment (proximal, middle and distal phalanges). The length of each finger segment is adapted by adding $c*(error_{average}-2)[cm]$, if $error_{average} \geq 2$, to the original finger segment length, where c equals 0.01, 0.03 and 0.1 for the proximal, middle and distal phalanges, respectively. These parameters were chosen manually to provide a smooth adaptation, and result in an added length of 0.14 cm per each degree of error above 2 deg (up to 8 cm when the average error magnitude for that finger is close to 60°). We only alter the length component of the finger (along the x-axis).

Pilot tests informed the specific mappings described above during development. Figure 3 illustrates these relationships. Feedback for all types of hand augmentations is provided in real-time.

4 Evaluation

We conducted a user study to understand how different elements of AR hand augmentations showing error feedback influence user's performance and preferences. Specifically, we aimed to answer the following research questions:

- RQ1: Can hand augmentations for error feedback, including manipulations in color, shape and placement, expedite the time needed to accurately replicate a demonstrated gesture?
- RQ2: Can hand augmentations for error feedback, including manipulations in color, shape and placement, improve the accuracy of gesture replication in a limited time period?

4.1 Experimental design

We use a within-subject design with two independent variables, FEEDBACK STYLE with three levels (color feedback, shape feedback, no explicit feedback), and PLACEMENT with two levels (adjacent, superimposed), resulting in a total of six conditions. We implemented color and shape feedback as described previously. We further included a no-explicit feedback condition as a baseline. For this condition, users only saw an AR hand demonstrating the target pose but did not receive any feedback on their own accuracy. As dependent variables, we measured the

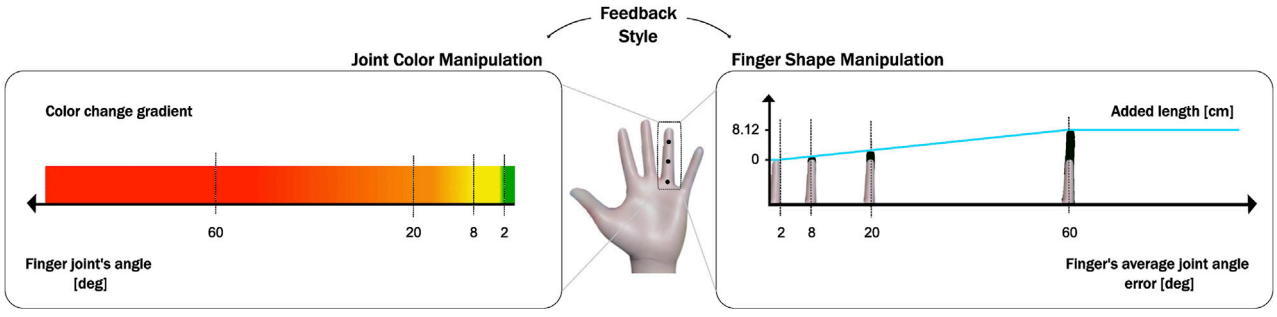

FIGURE 3
Side-by-side visual explanation of how angular error maps to two feedback types. The left panel shows a horizontal angular scale with key marks at 2°, 8°, 20°, and 60°, each mapped to a specific color in a gradient — green at 2°, yellow at 8°, orange at 20°, and red at 60° — representing the Color Feedback style. The right panel displays corresponding finger segments for each of those error values, with gradually increasing finger lengths matched to a linear function. The segment length increases up to a maximum of approximately 8.12 cm (keeping constant after that - from 60 degrees onwards), illustrating how Shape Feedback visually exaggerates the magnitude of error by elongating fingers proportionally.

TABLE 1 Questions included in the questionnaires to measure subjective task load [NASA TLX Hart and Staveland (1988)], hand augmentations usability [SUS Brooke et al. (1996) and Distraction], and preferences (semi-structured interview).

Questions	
Perceived Task Load	TLX1: How mentally demanding was this task?
	TLX2: How physically demanding was the task?
	TLX4: How successful do you think you were in accomplishing this task? (Note: here 1 is Perfect and 7 is Failure)
	TLX5: How hard did you have to work for accomplishing this task?
	TLX6: How insecure, discouraged, irritated, stressed and annoyed were you by performing this task?
Hand Augmentations' Understanding	SUS1: I think I would like to use this system frequently
	SUS2: I found the system unnecessarily complex
	SUS3: I thought the system was easy to use
	SUS9: I felt very confident using the system
	SUS10: I needed to learn a lot of things before I could get going with this system
	Distraction: I was distracted by the actions of the system
Interview	I1: What was the reasoning behind your preference scheme? Can you tell me about positive and negative points?
	I2: What improvements do you think we could do to our visualizations?
	I3: Do you have any ideas for cool visualizations besides what you experienced with using color and length?
	I4: What application scenarios would you see this being useful in?

time (recorded in milliseconds, reported in seconds) needed to replicate a gesture with a fixed target accuracy on our first task (RQ1), and the minimum average replication error in degrees reached during and after training over a fixed short period of time for the second task (RQ2). Additionally, we collected subjective ratings on perceived task load using a subset of questions from the NASA task load index (TLX) questionnaire (Hart and Staveland, 1988), users' ability to understand the hand augmentation using a subset of questions from the System Usability Scale (SUS) (Brooke et al., 1996), and distraction using an additional question. Questions were answered on a seven-point Likert scale, from Not at all (1) to Extremely (7) for the NASA TLX, or Strongly Disagree (1) to Strongly Agree (7) for the SUS. Our final post-task questionnaire can be found in Table 1. We note that we chose to exclude questions on time demand (TLX) and system-related usability questions (SUS) since these were not directly relevant to the individual augmentations and since including all questions would have significantly increased the duration of the study (already at 90 min). We did not compute the full SUS score since we are not evaluating a system.

Additionally, we conducted semi-structured post-experiment interviews to collect additional qualitative feedback. During the interviews, we showed participants representative illustrations of each of the six conditions they experienced, and asked them to rank these by preference, as well as inquiring about their reasoning behind preference.

4.2 Tasks

Participants were asked to perform two different tasks, which align with our two research questions on replication speed (RQ1) and replication accuracy (RQ2). The tasks are depicted in Figures 4, 5. For *Task 1*, the goal was for participants to replicate a set of consecutive different gestures as fast as possible. Participants were shown different gestures and had to reach an average joint accuracy of 2°. We chose a 2 deg error threshold based on pilot tests: when using 1.5 deg, pilot testers struggled to get the gestures correct; a 3 deg threshold led to gestures being marked as correct even though they were qualitatively different. We found 2 deg to balance difficulty and success rate. The error was calculated as the average of individual joint rotations.

For *Task 2*, the goal was to replicate a gesture as accurately as possible within a fixed time period of 10 s. Participants would repeatedly see individual gestures three times and were asked to replicate them to the best of their abilities. Between the trials, they took a 10-s break. After the third break, they were asked to replicate the same gesture from memory, i.e., without seeing the target gesture and without any feedback. We used a mix of gestures, including gestures from Portuguese Sign Language and Mudras Indian dance, in a total of 30 different randomized gestures (see Figure 6). Gestures were selected to include subtle variations that may appear similar but are distinct (i.e., there is no ambiguity in recognition). This ensured that participants could accurately perform different fine-grained gestures.

4.3 Procedure

Participants first completed the consent form and demographic questionnaires. Prior to the main task, participants were given a short presentation explaining each feedback style to ensure they understood how to interpret both color and shape cues. Participants then completed a

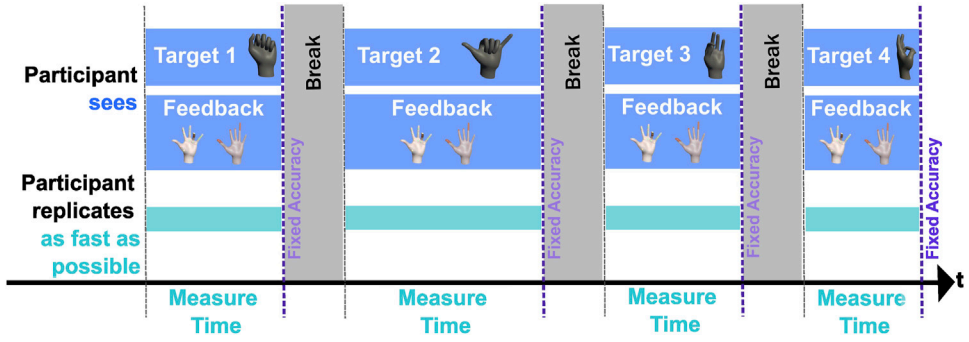

FIGURE 4
Timeline diagram illustrating the structure of Task 1, where participants replicate hand gestures as fast as possible. The process consists of four cycles. In each cycle, the participant first sees a target gesture (black hand image), and at the same time sees feedback (white hands with visible finger joints). Below, a blue bar indicates the participant replicates the gesture, trying to match it to a fixed accuracy threshold. After each successful replication (for corresponding duration), a Break period is shown in grey. This cycle repeats for four gestures. A horizontal black time arrow at the bottom shows progression through the experiment. Timing is adaptive: measurement ends once the participant reaches the required accuracy.

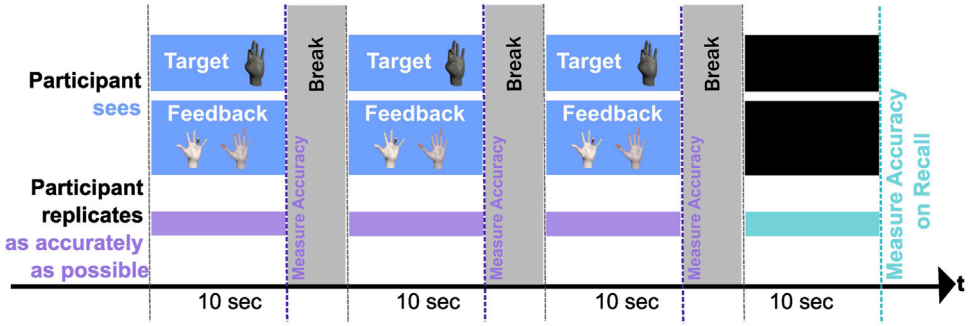

FIGURE 5
Timeline diagram illustrating the structure of Task 2, where participants replicate hand gestures as accurately as possible. Each trial is 10 seconds long. The participant sees a Target gesture (black hand image) and corresponding Feedback (white hand pair) during the first three trials. Each is followed by a Break period shown in grey. In the fourth trial, both the target and feedback are absent (black rectangles), and the participant must recall the gesture from memory. Accuracy is measured for each trial, with the final trial labeled "Measure Accuracy on Recall." The entire sequence is organized along a horizontal time axis with consistent 10-second intervals.

short tutorial in which they were introduced to each feedback condition, including shape feedback. During this familiarization phase, participants were able to freely interact with the system and ask questions to clarify any uncertainties. This introduction aimed to reduce confusion and support consistent interpretation across conditions. Afterwards, they performed Task 1 ("as fast as possible") under all six conditions, counterbalanced using a Latin square. The procedure is illustrated in Figure 4. At the beginning of the task, participants were instructed to place their hand within a specified area on the table, in a relaxed position. The same position was maintained during each break. After receiving initial instructions, they were presented with the first target gesture and prompted to replicate it. Upon successful replication at target accuracy (maximum average angular error of 2°), participants took a 10-s break. This process was repeated for three additional different gestures under the same condition, followed by the post-condition questionnaires. After a 1-min rest period, the entire task was repeated for five more conditions, each involving four additional different gestures.

Participants then took a 5-min break and continued with Task 2 ("as accurate as possible"). Participants' initial poses and instructions were similar. For each condition, participants had to replicate the same gesture as accurately as possible three times, each within a 10-s time interval. Afterwards, they replicated the gesture a fourth time without any visual guidance, followed by the post-condition questionnaire. Each condition included a different gesture.

After completing all conditions, we asked participants for their preference ranking and conducted the semi-structured interviews. The experiment took approximately 90 min per participant, and participants were compensated with a 30 Amazon gift card.

4.4 Participants and apparatus

We ran an *a priori* power analysis using G* Power 3.1 (Faul et al., 2009) to determine an appropriate sample size. We chose two effect sizes, f = 0.25 and f = 0.5, corresponding to small and medium effect

FIGURE 6
Complete list of gestures participants needed to perform during the experiment. All participants experienced the same 30 gestures: 24 during Task 1 (4 trials × 6 conditions) and 6 during Task 2, randomized across participants.

sizes, respectively, to determine the appropriate range for the sample size. We set an alpha error probability of $\alpha = 0.05$ and a power of $\beta = 0.8$. Since each condition consists of 1 or 4 trials (for *Replication Accuracy* and *Replication Time* respectively), we tested setting the number of measurements in G* Power from 6 to 24 (4 * 6 conditions). The number of groups was dependent upon the within-subject factors, which, in the case of our experiment, was 6. Finally, the correlation among repeated measures was left at the default value of 0.5. The power analysis revealed that we would need 12 participants to obtain a medium effect size. We also considered prior similar experiments [e.g., (Sodhi et al., 2012; Fidalgo et al., 2023; Schjerlund et al., 2021)], which had a similar number of participants.

We recruited 18 paid participants (10 male, 8 female), all students and research assistants from various fields of study or staff from a local university, with an average age of 26 years ($SD = 4.0$). Participants had different levels of experience with using both AR ($M = 2.5$, $SD = 1.15$) and VR ($M = 3.0$, $SD = 1.19$), and were not familiar with Portuguese Sign Language ($M = 1.1$, $SD = 0.24$) or Mudras Indian Dance movements ($M = 1.1$, $SD = 0.47$), on a scale from (1) low proficiency to (5) high proficiency. Based on self-reports, all participants had normal or corrected-to-normal vision, with 2 participants wearing contact lenses and 10 wearing glasses. None of the participants reported any color deficiencies; one participant was left-handed.

The study was conducted in a quiet experimental room. The AR scene was rendered using Unity 2020.3.14f1 and a Meta Quest Pro head-mounted display. The Meta Quest Pro headset was tethered to a desktop PC using Oculus Link to reduce latency and ensure a stable, high-performance setup. While we did not explicitly monitor tracking accuracy during the study, we conducted the experiment in a controlled, well-lit space with no visual clutter, and visually monitored hand tracking throughout. We also calibrated the physical space before each session to ensure hands were consistently recognized and accurately placed. These accuracy levels were deemed sufficient for the static hand pose replication task used in our experiment. Our apparatus ran on a Windows 10 PC with a 12th Gen Intel(R) Core(TM) i7-12700H processor and an NVIDIA RTX A1000 graphics card.

5 Results

We analyzed both the performance data and subjective measures using a series of 2 × 3, two-way repeated-measures ANOVAs with Aligned Rank Transformation (ART) applied Wobbrock et al. (2011), as data violated the normality assumption (Shapiro-Wilk test $p < .05$). The two within-subjects factors were Feedback Style (3 levels: None, Color, Shape) and Placement (2 levels: Superimposed, Adjacent). We apply Greenhouse-Geisser correction when the equal-variances assumption is violated (Mauchly's test $p < .05$). We tested for main effects of feedback style and placement, and conducted pairwise *post hoc* comparisons with Bonferroni adjustment *within that factor*, collapsed across the other, when a main effect was found. The statistical analysis was performed using the IBM SPSS software, and the ARTool for Windows Wobbrock et al. (2011). In the following, we report significant differences across the independent variables. Note that subjective preferences in Section 5.3 were analyzed separately using Friedman's ANOVA, treating the six feedback-placement combinations as a single 6-level factor, and applying 15 pairwise Wilcoxon signed-rank tests (Bonferroni corrected). All results are illustrated in Figure 7.

In summary, we found that shape feedback increases the time it takes to replicate a gesture compared to color feedback or no explicit feedback. Results also show that participants are faster to reach

10.3389/frvir.2025.1574965

FIGURE 7
Composite chart displaying results from hand gesture replication tasks, including replication time, accuracy, subjective preference, NASA TLX workload subscales, and system usability subscales, with comparisons across feedback types and placements. All plots use color and pattern coding: teal for color feedback, blue for shape, purple for no feedback, and striped bars pattern for adjacent vs circle pattern for superimposed placements.

target accuracy with superimposed feedback placement than with the target hands adjacent. Interestingly, subjective preferences show that participants preferred color feedback paired with adjacent placements, even though this resulted in slower replication speed. Additionally, neither feedback type nor placement affect replication accuracy on short-term recall.

10.3389/frvir.2025.1574965

TABLE 2 Summary of means and standard deviations for both tasks: Task 1 - Replication Time - and Task 2 - Replication Accuracy - both for short time recall and over trials.

Condition	Placement	T1: Replication time (seconds)		T2: Replication accuracy (degrees)			
		M	SD	Overall		Recall	
				M	SD	M	SD
Color	Superimposed	7.56	9.25	2.58	0.92	2.63	0.96
	Adjacent	7.87	9.30	2.81	1.17	2.88	1.70
Length	Superimposed	10.17	10.69	2.88	1.21	2.97	1.77
	Adjacent	23.05	29.31	3.19	1.29	3.39	1.46
None	Superimposed	6.97	7.64	2.67	0.71	2.71	1.10
	Adjacent	8.81	10.42	2.69	0.72	3.03	1.24

5.1 Performance

Performance data is illustrated in Figures 7.1,7.2, and summarized in Table 2. We replaced outliers further than two standard deviations from the global mean by the global mean for that condition. These corresponded to approximately 4% and 6% of the total collected data for replication time and accuracy (both for accuracy on recall and accuracy loss), respectively. We performed pairwise *post hoc* tests with a Bonferroni adjustment of $\alpha = 0.008$ (6 pairwise comparisons).

5.1.1 Task 1 - replication time

Results indicate a main effect on replication time for FEEDBACK ($F_{1.726,122.574} = 18.901$, $p < .001$, $\eta_p^2 = 0.21$), and PLACEMENT ($F_{1,71} = 25.173$, $p < .001$, $\eta_p^2 = 0.262$), and an interaction effect of both factors, FEEDBACK * PLACEMENT ($F_{2,142} = 7.748$, $p < .001$, $\eta_p^2 = 0.098$). Post-hoc tests showed that replication time was significantly lower for both color feedback ($M = 7.71$, $SD = 9.11$) and no explicit feedback ($M = 7.89$, $SD = 8.94$), when compared to shape feedback ($M = 16.61$, $SD = 22.86$), both $p < .05$. Having color feedback did not yield significantly different performance compared to having no explicit feedback during the task, while shape feedback actually increased the time needed to reach target accuracy. Post-hoc tests also revealed having the target hands in a superimposed placement ($M = 8.24$, $SD = 9.22$) significantly reduces replication time compared to adjacent placement ($M = 13.24$, $SD = 19.83$), $p < .001$. Accuracy was constant at 2°, as this was controlled for in the task.

5.1.2 Task 2 - replication accuracy

For the accuracy on short time recall reached on Task 2, statistical analysis did not indicate a main effect for either FEEDBACK ($p = 0.973$), or PLACEMENT ($p = 0.797$), nor for their interaction FEEDBACK*PLACEMENT ($p = 0.858$). Participants' average accuracy was comparable across all conditions ($M = 2.93$, $SD = 1.11$). Replication time was constant at 10 s, as this was controlled during the task.

Additionally, we evaluate the relative accuracy loss over conditions by comparing the maximum accuracy reached during the first 3 trials (where participants could see the target and the feedback) with the accuracy on the fourth trial (where participants replicate based on recall only). Statistical analysis did not indicate a main effect for either FEEDBACK ($p = 0.704$), or PLACEMENT ($p = 0.805$), nor for their interaction FEEDBACK*PLACEMENT ($p = 0.698$). The average relative accuracy loss from training to recall was comparable across all conditions (decreased accuracy by $M = 44.8\%$, $SD = 55.03\%$).

5.2 Subjective ratings

Results on the subjective scales for NASA TLX and SUS questionnaires, for both Task 1 and Task 2, are illustrated in Figures 7.4, 7.5, respectively. We performed pairwise *post hoc* tests with a Bonferroni adjustment of $\alpha = 0.008$ (6 pairwise comparisons).

5.2.1 Task 1 - replication speed

For the NASA TLX questions answered concerning Task 1 (replication speed), we observed no significant effect of either feedback or placement in either of the subscales.

Only the SUS subscale on learning (*"I needed to learn a lot of things before I could get going with this system"*) showed a main effect for feedback ($F_{2,34} = 3.642$, $p = 0.037$, $\eta_p^2 = 0.176$). However, *post hoc* tests did not show significant differences on pairwise comparisons.

5.2.2 Task 2 - replication accuracy

For Task 2 (replication accuracy), results on NASA TLX subscales revealed a statistically significant effect of feedback on *frustration*, ($F_{2,34} = 3.385$, $p = 0.046$, $\eta_p^2 = 0.166$). However, *post hoc* tests did not show significant differences on pairwise comparisons.

We also found a significant effect of both feedback ($F_{2,34} = 4.107$, $p = 0.025$, $\eta_p^2 = 0.195$) and PLACEMENT ($F_{1,17} = 5.214$, $p = 0.036$, $\eta_p^2 = 0.235$) on *physical demand*. Post-hoc tests showed that having shape feedback made the task physically less demanding ($M = 2.6$, $SD = 1.7$) than having no explicit feedback ($M = 2.4$, $SD = 1.7$), p = 0.016. Post hoc tests

also revealed that participants rated physical demand higher for conditions with superimposed placement ($M = 2.7$, $SD = 1.7$) than with adjacent placement ($M = 2.4$, $SD = 1.8$), p = 0.008. Having both the target hands superimposed to the user's hands and shape feedback made the task of training to replicate a gesture based on recall physically more demanding. Nevertheless, small differences might make them negligible in practice.

We also found an interaction effect FEEDBACK*PLACEMENT on *mental demand* ($F_{2,30} = 4.233$, $p = 0.024$, $\eta_p^2 = 0.220$). While none of the posthoc tests revealed statistically reliable differences, this might indicate that certain combinations of feedback and placement could induce higher cognitive load than others.

For the SUS questions, similarly to Task 1, we only found a significant effect of FEEDBACK on the SUS subscale for learning. ($F_{2,34} = 3.381$, $p = 0.046$, $\eta_p^2 = 0.166$). Post hoc tests revealed that participants rated higher in how much they needed to learn before they could get going with the task for conditions where they had shape feedback, compared to receiving no feedback, p = 0.011.

5.3 Subjective feedback

To complement our quantitative findings, we conducted semi-structured post-experiment interviews to collect additional qualitative feedback. During the interviews, participants were shown representative illustrations of each of the six feedback-placement conditions they experienced. They were asked to rank these combinations by preference and explain their reasoning, including perceived benefits, drawbacks, potential improvements, and application scenarios for the feedback mechanisms experienced.

We leverage these qualitative insights to contextualize the quantitative results reported in Section 5. For example, while color feedback with adjacent placement was statistically the most preferred condition, participants' explanations provide insight into why this was the case, highlighting the perceived intuitiveness and clarity of color cues. On the other hand, shape feedback's low rankings are supported by participant comments describing confusion and lack of directional guidance. This alignment supports the validity of our findings and offers design-relevant nuance beyond statistical significance.

We gathered participant reasoning to supplement and interpret the quantitative results. The comments bellow, though not exhaustive, highlight recurring user experiences that can inform design considerations and future work; we use $N = X$ to indicate how many participants expressed similar remarks.

5.3.1 Preferences

We asked participants to think about their preferences over the whole experiment and perform a sorting task. Most participants (N = 12) preferred color feedback paired with adjacent placement as their favourite, and shape feedback with superimposed placement as least favourite (N = 10). We treated participants' general preferences as a single 6-level factor (six different feedback-placement combinations). Statistical analysis with a Friedman's ANOVA showed significant differences in the ranking of participants' general preferences, $X^2(5) = 25.655$, $p < .001$). Post hoc analysis with Wilcoxon signed-rank Tests with Bonferroni correction (adjusted $\alpha = 0.0033$ to account for 15 pairwise comparisons) revealed color feedback paired with adjacent placement was preferred when compared to color feedback paired with superimposed placement, $p = 0.003$, shape feedback paired with superimposed and adjacent placement, $p < .001$ and $p = 0.001$ respectively, and to having to no explicit feedback with adjacent placement, $p = 0.001$. Figure 7.3 illustrates participants' subjective ranking scores for order of preference for all conditions (1-6 from best to worst).

5.3.2 Feedback type

When asked to explain their reasoning for ranking preferences, participants mentioned that the colored joints were "useful" ($N = 5$) and "intuitive" ($N = 3$). Nevertheless, some participants ($N = 3$) mentioned that the colored joint feedback could be distracting, making them focus more on getting all the joints correct (green) instead of focusing on the gesture itself (P6: "When feedback was removed [I] did not remember [the] actual gesture anymore"). Most participants justified their least favorite choice by describing the shape feedback as "confusing" ($N = 7$) and "not useful" ($N = 3$). It did not indicate which direction to move to reach the correct position, so participants needed to proceed on a trial-and-error basis ($N = 4$).

5.3.3 Placement

Most participants expressed their preference for having the hands adjacent to theirs ($N = 10$), mentioning how it enabled them to see their hand clearly, using the adjacent target as a reference and the feedback for corrections (P18: "We want to remember, so colored joints really help a lot - if fingers drift you see the joint changing and come back to previous position. In this case, [it is] better to have an adjacent target because you want to see the feedback. Same for shape, useful to see the target hand as high level and then use feedback for drifting.") However, some participants preferred a superimposed placement, stating it was easier to follow ($N = 4$), and mentioned how they could quickly and intuitively replicate the base pose by looking at the superimposition of the target on their hand, and then use the feedback for smaller corrections (P9, P18), without their eyes needing to go back and forth (P8).

P18 expressed how having a superimposed target hand only, with no feedback, did not give enough confidence on how correct the movement was, but that they "liked the simplicity of the implicit feedback." Some participants also expressed how the combination of length extension or color with superimposed placement could be overwhelming, as there was "too much happening" ($N = 7$).

5.3.4 Improvements

Participants suggested applying color to the whole hand or selected zones such as individual fingers instead of joints ($N = 8$). Most participants also mentioned that they would have liked to see color integrated with length ($N = 7$), with most mentioning how having the finger error bars always red gave them a constant negative impression ($N = 9$). It would have been nice to include some more explicit cues on which direction they should move (P4, P16). Participants also suggested the inclusion of sound effects (P13), haptic feedback (P8), multiple perspectives of the user's own hands (P8, P13), or integrating more gamification ("User is gradually rewarded for correct behaviour: Color the full finger or parts of the hand when mostly right, kind of level unlock," P7).

5.3.5 Potential applications

Participants mentioned that they would find gestural feedback useful for applications such as teaching physical tasks where dexterity matters (N = 3), mentioning domains such as music e.g., learning to play the piano (P4, P5), playing the guitar (P11), or music conducting (P13); entertainment (puppet shows, P16) and shadow play (P16, P18), and in training workers on how to handle machinery in industrial settings (P15). Participants also mentioned its usefulness for learning language-based gestures (N = 6), although our profile questionnaire asking if they were familiar with Sign Language might have prompted this.

P8 and P9 also mentioned that such approaches could be useful in the context of interaction with new technologies, such as XR Gaming ("[In AR/VR] this could be part of the game instructions, when we need especial gestures," P8 and "the new Apple Vision Pro for learning how to control the headset," P9).

Finally, some participants mentioned extending a similar approach to full-body movements and posture feedback would be interesting. The context of sports (P3), dance and choreography practice (N = 3), and physical therapy and rehabilitation (N = 6) were mentioned, with P17 pointing out that this could "Reduce the need for one-on-one therapy". Participants also mentioned the usefulness in general medical applications and surgery (N = 4).

6 Discussion

We investigate the impact of different feedback forms and placements on users' performance and subjective experiences in a gesture replication task.

We observed that participants performed significantly faster when provided with color or no explicit feedback compared to shape feedback (RQ1). This suggests that while color cues or the absence of explicit feedback may facilitate faster task completion, shape feedback seems to introduce cognitive load or uncertainty, thus impeding performance. This is also reflected in participants' comments expressing frustration and confusion.

Furthermore, our results demonstrate that the placement of target hands relative to the user's hands also impacts replication time. Participants replicated gestures more quickly when the target hands were superimposed rather than adjacent to their own hands. This finding suggests that superimposed placement may offer perceptual advantages or facilitate spatial coordination, enabling users to align their movements with the target more efficiently.

Interestingly, the preference for superimposed placement does not align with subjective interview preferences. While some participants preferred this configuration due to its perceived ease of use, most highlighted the benefits of adjacent placement for providing clear visual reference points and facilitating feedback interpretation. This discrepancy might be because an adjacent placement mimics the real-life paradigm for seeing a demonstration, which creates a sense of familiarity that influences preference. The fact that participants were significantly faster (38%) when seeing the target gesture superimposed on their hand, but still preferred the adjacent placement, which underscores the complexity of introducing new interaction paradigms and how familiarity plays a role in user experience.

This divergence between performance metrics and subjective preferences highlights a critical tension in interface design: optimizing for efficiency does not always align with what users find intuitive or comfortable. Participants' preference for adjacent placement, despite its lower performance, may be shaped by perceptual familiarity and reduced cognitive demand during interpretation. The adjacent configuration aligns with established interaction metaphors - such as observing a demonstrator beside oneself - which may feel more natural, particularly for novice users. In contrast, the superimposed view, while more efficient, may demand greater perceptual adaptation. As an example, Van Beurden et al. (2011) found that while device-based interfaces scored higher on perceived performance, gesture-based interfaces were preferred for their hedonic qualities and fun. Indeed, prior work on gesture-based interfaces (Norman, 2010) suggests that learning curves and usability trade-offs are central in shaping user preferences, particularly when novel interaction paradigms are introduced. Future studies could explore how these preferences evolve over time and how training or increased exposure may shift the balance between efficiency and user comfort, and whether indirect mappings may offer retention or transfer benefits in more complex gesture learning tasks.

Additionally, while our study did not directly measure cognitive load beyond the NASA-TLX mental demand subscale, we did observe an interaction effect between feedback type and placement on reported mental effort. Although *post hoc* comparisons were not statistically significant, this finding may indicate that certain combinations of feedback and placement introduce greater cognitive burden. One possible explanation is that shape-based feedback, particularly in adjacent placements, may require more visual interpretation and impose higher attentional demand due to the spatial separation between the user's hand and the provided feedback. This aligns with prior work suggesting that spatially incongruent or complex visual feedback can increase cognitive processing load in AR/VR environments (Makransky et al., 2019; Cañas et al., 2005). As we did not design our study to isolate cognitive mechanisms, these results need to be interpreted cautiously. However, future research could more systematically examine the cognitive cost of different feedback strategies using complementary methods such as eye tracking, dual-task paradigms, or physiological measures (e.g., EEG, pupillometry).

Furthermore, our study did not find significant effects of feedback type or placement on replication accuracy (RQ2), suggesting that these factors may not strongly influence the fidelity of gesture replication in short-term recall tasks. This finding implies that while feedback and placement strategies may impact task efficiency and user experience, they may not necessarily affect the quality of task performance.

While our study focuses on short-term gesture replication as a proxy for early-stage learning, the long-term implications of continuous feedback need further exploration. Over-reliance on visual guidance may hinder the development of autonomous gesture performance, especially if users become dependent on external cues. This concern echoes findings in the AR literature, where persistent visual overlays have been shown to narrow attention and reduce awareness of physical context—a phenomenon known as attentional tunneling (Tang et al., 2003; Syiem et al., 2021). Some participants in our study also described continuous feedback as distracting, suggesting that more adaptive or phased feedback strategies may better support learning. Techniques

such as faded feedback (Goodman and Wood, 2009) or error-based scaffolding (Finn and Metcalfe, 2010), explored in different contexts, may encourage gradual internalization of gestures and improve long-term retention. Future work should investigate how gesture training systems can balance immediate guidance with long-term skill independence.

6.1 Design guidelines

We offer preliminary guidelines for designing effective AR tutorials based on the previous discussion. These are informed by the observed findings, though we caution that user preferences and task contexts may influence their applicability.

6.1.1 Color feedback

Color feedback, which uses color transitions (e.g., red to green) to signal gesture angular correctness, can help users quickly detect and respond to errors. It was associated with faster task completion in our study, suggesting it may be well-suited for tasks that demand quick, responsive correction. However, its effectiveness may depend on users' familiarity with such visual encoding.

6.1.2 Shape feedback

Shape exaggeration may help visualize error magnitude, particularly for more ambiguous gestures. However, participants in our study found it less efficient, potentially due to the cognitive effort required to interpret the exaggerated shapes. Clear directional cues are necessary to mitigate this issue and improve interpretability.

6.1.3 Placement

Adjacent placement was preferred for its clarity, supporting users to easily compare their movements with the intended gestures without obscuring their view. Alternatively, superimposed placement led to faster performance but was not as well-liked, indicating a trade-off between efficiency and user comfort. These preferences suggest placement should be adapted to user needs and task requirements, potentially being more effective for tasks requiring high precision and rapid skill acquisition.

6.1.4 Ergonomics

The study findings emphasized that while feedback type and placement significantly speed up task completion, they do not compromise the accuracy of gesture replication. Despite the operational benefits of superimposed feedback, many users prefer adjacent placement due to its straightforward nature, which underscores the importance of aligning tutorial design with user comfort to enhance efficiency and engagement. Designers should consider ergonomic and perceptual load when implementing these cues, especially for sustained use in learning environments.

6.1.5 Complexity trade-offs

Our findings suggest that balancing between adjacent and superimposed feedback based on task complexity and user experience can lead to more effective and satisfying educational experiences in AR settings. We recommend designers balance clarity, efficiency, and user preferences when selecting feedback and placement modalities in order to optimize learning outcomes.

6.2 Limitations and future work

6.2.1 Design space

When creating our design space, we chose *Color* and *Placement* as factors, as these are commonly used to provide error feedback for data visualizations and 2D interfaces (e.g., color: heatmaps; placement: error indicators next to UIs); and *Shape* to cover for beyond-real interactions, as typical in XR environments [e.g., shape: body part elongation McIntosh et al. (2020); Poupyrev et al. (1996)]. We chose these augmentations as a starting point for our investigation due to the ubiquity of their 2D counterparts. We acknowledge, however, that the space of *possible* augmentations is significantly larger, and could include other factors such as motion paths, speeds, sound alerts, and other sensory feedback (haptics: texture, pressure, temperature).

6.2.2 Long-term learning

Our work currently focuses on improving gesture replication through a static hand pose replication task, which we use as a proxy for the initial stages of learning during a gesture-based tutorial. While our study suggests that feedback and placement may not necessarily impact the overall quality of task performance on short-term recall, results might be different when considering long-term learning. Future work could investigate whether hand augmentations may inadvertently lead to users over-relying on this form of guidance in the long term, hindering their autonomy in performing tasks when continuous visual cues are unavailable. Additionally, as some participants noted during the interviews, continuous feedback can be distracting, removing focus from the actual gesture replication. Hence, we should carefully consider the potential impact of continuous and dynamic feedback on overall learning retention.

6.2.3 Static vs. dynamic gestures

We focus on replicating a static hand pose in terms of the accuracy of the hand pose itself. We believe that those cover a large space of tutorials and applications. However, many actions exist that rely on dynamics and hand movement. We hope to expand our work to those areas, for example, by providing feedback on motion paths and speed.

6.2.4 Tracking

Our motion capture method relies on the Quest Pro's in-built hand tracking, which is unreliable when gestures have intricate poses with obstructed fingers. In our study, this mostly led to challenges when participants noticed that their virtual hands did not match their actual hands perfectly. This problem makes generating accurate automatic feedback even more difficult, especially when considering more complex tasks that include hand-object interactions, where full portions of the hand might be occluded. We hope to leverage higher-accuracy marker-based motion capture systems in the future.

6.2.5 Multi-modal feedback

Finally, our approach focuses primarily on visual cues. Future work could explore integrating other modalities, such as audio and haptic sensations (Schütz et al., 2022; Cho et al., 2024), including pressure, texture, and temperature changes. For example, texture

and temperature changes could be particularly relevant in surgical training.

7 Conclusion

We explored different automatic feedback mechanisms through hand augmentations to enhance gesture-based tutorials in AR environments. Through a user study with 18 participants, we evaluated different styles of hand augmentations, combining different types of feedback at different placements. We observed that participants performed significantly faster when provided with color or no explicit feedback than shape feedback, indicating the potential cognitive burden introduced by manipulating shape. Our study did not find significant effects of feedback type or placement effects on replication accuracy, suggesting that while these factors may influence task efficiency, they may not strongly affect overall task proficiency in the short term. Interestingly, while having the target hands in a superimposed placement significantly reduces replication time, most participants preferred adjacent placement, highlighting its clarity and facilitation of feedback interpretation. In light of our findings, we are optimistic that advancing hand augmentation technologies in AR will significantly streamline user interactions and enhance learning outcomes, paving the way for more natural and effective learning virtual environments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, upon request to the corresponding author.

Ethics statement

The studies involving humans were approved by Carnegie Mellon University IRB Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

CF: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. YY: Software, Writing – original draft, Writing – review and editing. MS: Conceptualization, Writing – original draft, Writing – review and editing. JJ: Methodology, Supervision, Writing – original draft, Writing – review and editing. DL: Conceptualization, Formal Analysis, Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is co-financed by Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) and the Carnegie Mellon Portugal Program fellowship SFRH/BD/151465/2021.

Acknowledgments

The authors would like to thank the participants who participated in the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. I acknowledge the use of GPT4 (OpenAI, https://chatgpt.com) at the drafting stage of paper writing, and proofreading of the final draft.

Publisher's note

References

Abtahi, P., Hough, S. Q., Landay, J. A., and Follmer, S. (2022). "Beyond being real: a sensorimotor control perspective on interactions in virtual reality," in *Proceedings of the 2022 CHI conference on human factors in computing systems*, 1–17.

Amores, J., Benavides, X., and Maes, P. (2015). "Showme: a remote collaboration system that supports immersive gestural communication," in *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 15, 1343–1348. doi:10.1145/2702613.2732927

Barioni, R. R., Costa, W., Aleluia, A., and Teichrieb, V. (2019). "Balletvr: a virtual reality system for ballet arm positions training," in *2019 21st symposium on virtual and augmented reality (SVR)* (IEEE), 10–16.

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4–7.

Brunet, P., and Andújar, C. (2015). Immersive data comprehension: visualizing uncertainty in measurable models. *Front. Robotics AI* 2, 22. doi:10.3389/frobt.2015.00022

Büttner, S., Prilla, M., and Röcker, C. (2020). "Augmented reality training for industrial assembly work - are projection-based ar assistive systems an appropriate tool for assembly training?," in *Proceedings of the 2020 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 20, 1–12. doi:10.1145/3313831.3376720

Cañas, J. J., Quesada, J. F., Antolí, A., and Fajardo, I. (2005). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics* 48, 1283–1298.

Cho, H., Sendhilnathan, N., Nebeling, M., Wang, T., Padmanabhan, P., Browder, J., et al. (2024). "Sonohaptics: an audio-haptic cursor for gaze-based object selection in xr," in *Proceedings of the 37th annual ACM symposium on user interface software and technology* (New York, NY, USA: Association for Computing Machinery). UIST '24. doi:10.1145/3654777.3676384

Clarke, C., Cavdir, D., Chiu, P., Denoue, L., and Kimber, D. (2020). "Reactive video: adaptive video playback based on user motion for supporting physical activity," in *Proceedings of the 33rd annual ACM symposium on user interface software and technology*, 196–208.

de Koning, B. B., Hoogerheide, V., and Boucheix, J.-M. (2018). Developments and trends in learning with instructional video. *Comput. Hum. Behav.* 89, 395–398. doi:10.1016/j.chb.2018.08.055

Diller, F., Scheuermann, G., and Wiebel, A. (2024). Visual cue based corrective feedback for motor skill training in mixed reality: a survey. *IEEE Trans. Vis. Comput. Graph.* 30, 3121–3134. doi:10.1109/TVCG.2022.3227999

Dürr, M., Weber, R., Pfeil, U., and Reiterer, H. (2020). "Eguide: investigating different visual appearances and guidance techniques for egocentric guidance visualizations," in *Proceedings of the fourteenth international conference on tangible, embedded, and embodied interaction*, 311–322.

Endow, S., and Torres, C. (2021). ""i'm better off on my own": understanding how a tutorial's medium affects physical skill development," in *Designing interactive systems conference 2021* (New York, NY, USA: Association for Computing Machinery), 21, 1313–1323. doi:10.1145/3461778.3462066

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: tests for correlation and regression analyses. *Behav. Res. methods* 41, 1149–1160. doi:10.3758/brm.41.4.1149

Feuchtner, T., and Müller, J. (2018). "Ownershift: facilitating overhead interaction in virtual reality with an ownership-preserving hand space shift," in *Proceedings of the 31st annual ACM symposium on user interface software and technology*, 31–43.

Fidalgo, C., Yan, Y., Cho, H., Sousa, M., Lindlbauer, D., and Jorge, J. (2023a). *A survey on remote assistance and training in mixed reality environments*. IEEE VR. doi:10.1109/TVCG.2023.3247081

Fidalgo, C. G., Sousa, M., Mendes, D., Dos Anjos, R. K., Medeiros, D., Singh, K., et al. (2023b). "Magic: manipulating avatars and gestures to improve remote collaboration," in *2023 IEEE conference virtual reality and 3D user interfaces (VR)*, 438–448. doi:10.1109/VR55154.2023.00059

Finn, B., and Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Mem. and cognition* 38, 951–961. doi:10.3758/MC.38.7.951

Flanagan, J. R., and Johansson, R. S. (2002). Hand movements. *Encycl. Hum. brain* 2, 399–414. doi:10.1016/b0-12-227210-2/00157-6

Goodman, J. S., and Wood, R. E. (2009). Faded versus increasing feedback, task variability trajectories, and transfer of training. *Hum. Perform.* 22, 64–85. doi:10.1080/08959280802541013

Goto, M., Uematsu, Y., Saito, H., Senda, S., and Iketani, A. (2010). "Task support system by displaying instructional video onto ar workspace," in *2010 IEEE international symposium on mixed and augmented reality*, 83–90. doi:10.1109/ISMAR.2010.5643554

Gupta, A., Fox, D., Curless, B., and Cohen, M. (2012). "Duplotrack: a real-time system for authoring and guiding duplo block assembly," in *Proceedings of the 25th annual ACM symposium on user interface software and technology* (New York, NY, USA: Association for Computing Machinery), 389–402. doi:10.1145/2380116.2380167

Hart, S. G., and Staveland, L. E. (1988). "Development of nasa-tlx (task load index): results of empirical and theoretical research," in *North-holland, vol. 52 of* Advances in psychology. Editors H. Mental Workload, P. A. Hancock, and N. Meshkati, 139–183. doi:10.1016/S0166-4115(08)62386-9

Herbert, B., Ens, B., Weerasinghe, A., Billinghurst, M., and Wigley, G. (2018). Design considerations for combining augmented reality with intelligent tutors. *Comput. and Graph.* 77, 166–182. doi:10.1016/j.cag.2018.09.017

Huang, G., Qian, X., Wang, T., Patel, F., Sreeram, M., Cao, Y., et al. (2021). "Adaptutar: an adaptive tutoring system for machine tasks in augmented reality," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, 1–15.

Jacobs, K. W., and Suess, J. F. (1975). Effects of four psychological primary colors on anxiety state. *Percept. Mot. Ski.* 41, 207–210. doi:10.2466/pms.1975.41.1.207

Jamet, E., Gavota, M., and Quaireau, C. (2008). Attention guiding in multimedia learning. *Learn. Instr.* 18, 135–145. doi:10.1016/j.learninstruc.2007.01.011

Jo, H.-Y., Seidel, L., Pahud, M., Sinclair, M., and Bianchi, A. (2023). "Flowar: how different augmented reality visualizations of online fitness videos support flow for at-home yoga exercises," in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1–17.

Knierim, P., Schwind, V., Feit, A. M., Nieuwenhuizen, F., and Henze, N. (2018). "Physical keyboards in virtual reality: analysis of typing performance and effects of avatar hands," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–9.

Krolovitsch, A.-K., and Nilsson, L. (2009). 3d visualization for model comprehension a case study conducted at ericsson ab. B.S. thesis

Langlotz, T., Zingerle, M., Grasset, R., Kaufmann, H., and Reitmayr, G. (2012). "Ar record&replay: situated compositing of video content in mobile augmented reality," in *Proceedings of the 24th Australian computer-human interaction conference* (New York, NY, USA: Association for Computing Machinery), 318–326. doi:10.1145/2414536.2414588

Li, J., Sousa, M., Mahadevan, K., Wang, B., Aoyagui, P. A., Yu, N., et al. (2023). "Stargazer: an interactive camera robot for capturing how-to videos based on subtle instructor cues," in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1–16.

Lilija, K., Kyllingsbæk, S., and Hornbæk, K. (2021). "Correction of avatar hand movements supports learning of a motor skill," in *2021 IEEE virtual reality and 3D user interfaces (VR)* (IEEE), 1–8.

Liu, R., Wu, E., Liao, C.-C., Nishioka, H., Furuya, S., and Koike, H. (2023a). "Pianohandsync: an alignment-based hand pose discrepancy visualization system for piano learning," in *Extended abstracts of the 2023 CHI conference on human factors in computing systems*, 1–7.

Liu, Z., Zhu, Z., Jiang, E., Huang, F., Villanueva, A. M., Qian, X., et al. (2023b). "Instrumentar: auto-generation of augmented reality tutorials for operating digital instruments through recording embodied demonstration," in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1–17.

Makransky, G., Terkildsen, T. S., and Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* 60, 225–236. doi:10.1016/j.learninstruc.2017.12.007

Mayer, R. E., Fiorella, L., and Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educ. Technol. Res. Dev.* 68, 837–852. doi:10.1007/s11423-020-09749-6

McIntosh, J., Zajac, H. D., Stefan, A. N., Bergström, J., and Hornbæk, K. (2020). "Iteratively adapting avatars using task-integrated optimisation," in *Proceedings of the 33rd annual ACM symposium on user interface software and technology*, 709–721.

Mohr, P., Mandl, D., Tatzgern, M., Veas, E., Schmalstieg, D., and Kalkofen, D. (2017). "Retargeting video tutorials showing tools with surface contact to augmented reality," in *Proceedings of the 2017 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 17, 6547–6558. doi:10.1145/3025453.3025688

Morillo, P., Garcia-Garcia, I., Orduna, J. M., Fernandez, M., and Juan, M. C. (2020). Comparative study of ar versus video tutorials for minor maintenance operations. *Multimedia Tools Appl.* 79, 7073–7100. doi:10.1007/s11042-019-08437-9

Norman, D. A. (2010). Natural user interfaces are not natural. *interactions* 17, 6–10. doi:10.1145/1744161.1744163

Oshita, M., Inao, T., Ineno, S., Mukai, T., and Kuriyama, S. (2019). Development and evaluation of a self-training system for tennis shots with motion feature assessment and visualization. *Vis. Comput.* 35, 1517–1529. doi:10.1007/s00371-019-01662-1

Ozcelik, E., Arslan-Ari, I., and Cagiltay, K. (2010). Why does signaling enhance multimedia learning? evidence from eye movements. *Comput. Hum. Behav.* 26, 110–117. doi:10.1016/j.chb.2009.09.001

Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. (1996). "The go-go interaction technique: non-linear mapping for direct manipulation in vr," in *Proceedings of the 9th annual ACM symposium on User interface software and technology*, 79–80.

Rajaram, S., and Nebeling, M. (2022). "Paper trail: an immersive authoring system for augmented reality instructional experiences," in *Proceedings of the 2022 CHI conference on human factors in computing systems*, 1–16.

Ricca, A., Chellali, A., and Otmane, S. (2020). "Influence of hand visualization on tool-based motor skills training in an immersive vr simulator," in *2020 IEEE international symposium on mixed and augmented reality (ISMAR)* (IEEE), 260–268.

Ricca, A., Chellali, A., and Otrnane, S. (2021). "The influence of hand visualization in tool-based motor-skills training, a longitudinal study," in *2021 IEEE virtual reality and 3D user interfaces (VR)* (IEEE), 103–112.

Schjerlund, J., Hornbæk, K., and Bergström, J. (2021). "Ninja hands: using many hands to improve target selection in vr," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, 1–14.

Schütz, L., Weber, E., Niu, W., Daniel, B., McNab, J., Navab, N., et al. (2022). enAudiovisual augmentation for coil positioning in transcranial magnetic stimulation. *Comput. methods biomechanics Biomed. Eng. Imaging and Vis.*, 1–8. doi:10.1080/21681163.2022.2154277

Skreinig, L. R., Stanescu, A., Mori, S., Heyen, F., Mohr, P., Sedlmair, M., et al. (2022). "Ar hero: generating interactive augmented reality guitar tutorials," in *2022 IEEE*

conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), 395–401. doi:10.1109/VRW55335.2022.00086

Sodhi, R., Benko, H., and Wilson, A. (2012). Lightguide: projected visualizations for hand movement guidance. *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 179–188. doi:10.1145/2207676.2207702

Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., and Maciocci, G. (2013). "Bethere: 3d mobile collaboration with spatial input," in *Proceedings of the SIGCHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 13, 179–188. doi:10.1145/2470654.2470679

Spielberger, C. D. (1970). Manual for the state-trait anxiety inventory (self-evaluation questionnaire). *(No Title)*

Stanescu, A., Mohr, P., Kozinski, M., Mori, S., Schmalstieg, D., and Kalkofen, D. (2023). "State-aware configuration detection for augmented reality step-by-step tutorials," in *2023 IEEE international symposium on mixed and augmented reality (ISMAR)* (IEEE), 157–166.

Syiem, B. V., Kelly, R. M., Goncalves, J., Velloso, E., and Dingler, T. (2021). "Impact of task on attentional tunneling in handheld augmented reality," in *Proceedings of the 2021 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3411764.3445580

Tang, A., Owen, C., Biocca, F., and Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. *Proc. SIGCHI Conf. Hum. factors Comput. Syst.*, 73–80. doi:10.1145/642611.642626

Tecchia, F., Alem, L., and Huang, W. (2012). "3d helping hands: a gesture based mr system for remote collaboration," in *Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry* (New York, NY, USA: Association for Computing Machinery), 323–328. doi:10.1145/2407516.2407590

Torres, C., and Figueroa, P. (2018). "Learning how to play a guitar with the hololens: a case study," in *2018 XLIV Latin American computer conference (CLEI)* (IEEE), 606–611.

Van Beurden, M. H., Ijsselsteijn, W. A., and de Kort, Y. A. (2011). "User experience of gesture based interfaces: a comparison with traditional interaction methods on pragmatic and hedonic qualities," in *International gesture workshop* (Springer), 36–47.

Voisard, L., Hatira, A., Sarac, M., Kersten-Oertel, M., and Batmaz, A. U. (2023). "Effects of opaque, transparent and invisible hand visualization styles on motor dexterity in a virtual reality based purdue pegboard test," in *2023 IEEE international symposium on mixed and augmented reality (ISMAR)* (IEEE), 723–731.

Wang, X., Lafreniere, B., and Zhao, J. (2024). "Exploring visualizations for precisely guiding bare hand gestures in virtual reality," in *Proceedings of the CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3613904.3642935

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 11, 143–146. doi:10.1145/1978942.1978963

Yamaguchi, M., Mori, S., Mohr, P., Tatzgern, M., Stanescu, A., Saito, H., et al. (2020). "Video-annotated augmented reality assembly tutorials," in *Proceedings of the 33rd annual ACM symposium on user interface software and technology* (New York, NY, USA: Association for Computing Machinery), 1010–1022. doi:10.1145/3379337.3415819

Yu, X., Angerbauer, K., Mohr, P., Kalkofen, D., and Sedlmair, M. (2020). "Perspective matters: design implications for motion guidance in mixed reality," in *2020 IEEE international symposium on mixed and augmented reality (ISMAR)* (IEEE), 577–587.

Yu, X., Lee, B., and Sedlmair, M. (2024). "Design space of visual feedforward and corrective feedback in xr-based motion guidance systems," in *Proceedings of the CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3613904.3642143

Zagermann, J., Pfeil, U., Fink, D., von Bauer, P., and Reiterer, H. (2017). Memory in motion: the influence of gesture- and touch-based input modalities on spatial memory. *Proc. 2017 CHI Conf. Hum. Factors Comput. Syst.* 17, 1899–1910. doi:10.1145/3025453.3026001

Zhou, Q., Irlitti, A., Yu, D., Goncalves, J., and Velloso, E. (2022). "Movement guidance using a mixed reality mirror," in *Proceedings of the 2022 ACM designing interactive systems conference*, 821–834.